Estimating One-Parameter Airport Arrival Capacity Distributions Using Stochastic Modeling

Tasha R. Inniss,

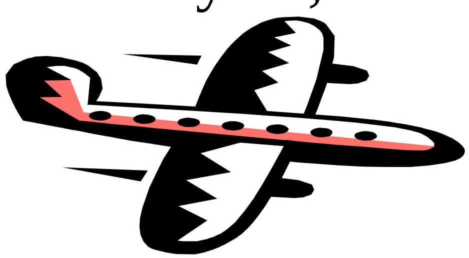
Clare Boothe Luce Professor of Mathematics,

Trinity College

Computer Science Research Institute-Seminar Series

Sandia National Laboratories

May 22, 2001



Background

- Method is needed to efficiently address *capacity-demand imbalances*
- To address and manage these imbalances, ATCSCC may institute *ground delay programs* (GDPs)
- Determining the amount of delay to assign in a GDP is known as the *ground holding problem* (GHP)
- GDP planning has become more efficient under a new collaborative process known as *Collaborative Decision Making* (CDM)

Goal/Motivation

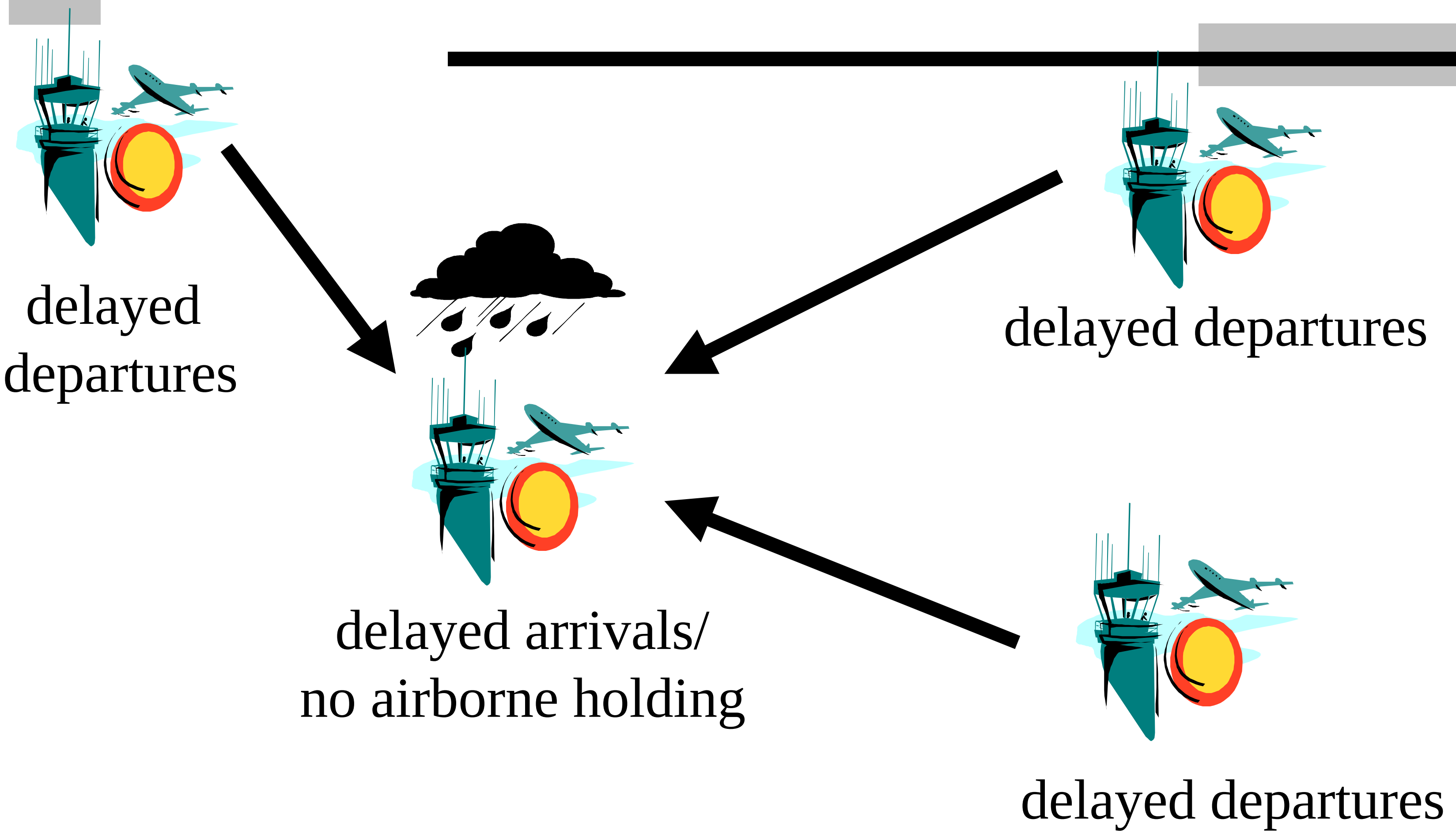
- Goal: Estimate airport arrival capacity distributions during inclement weather conditions

- Why?
  - Bad weather reduces capacity below demand
  - Implicit relationship between weather and capacity
  - Stochastic nature of weather makes it difficult to deterministically predict capacity
  - Required input to a class of stochastic ground holding models

Outline

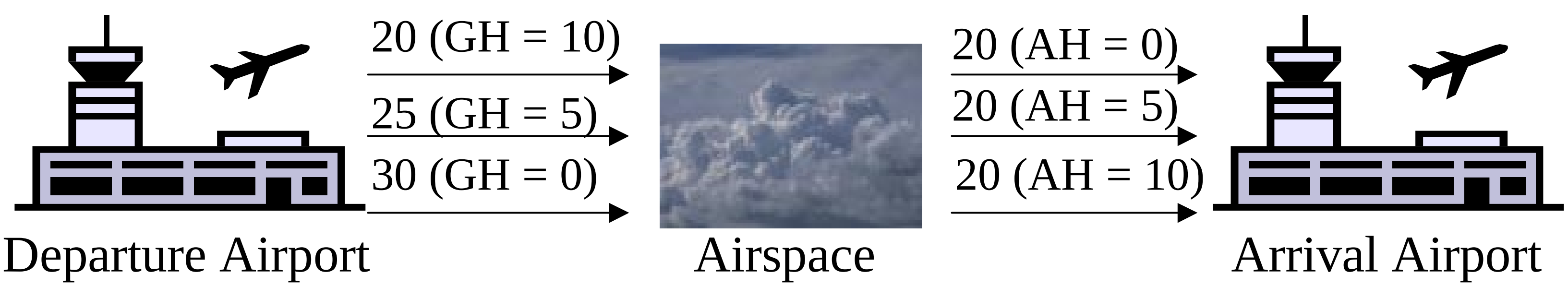
- Background
  - Discussion of a Ground Delay Program
  - Hoffman-Rifkin Static Stochastic Ground Holding (H-R) Model
- Capacity Scenarios (Arrival Capacity Distributions)
  - Conceptual Representation of ACDs
  - Generating overall distribution of ACDs
- Deriving Seasonal Distributions via “Seasonal Clustering”
- Adjusting assigned delay in dynamic GDPs
- Comparing results of H-R Model to Command Center Plans
- Conclusions/Future Work

What is a Ground Delay Program (GDP)?



Ground vs. Airborne Delay

- In a GDP, determining the optimal amount of ground delay to assign is known as the **Ground Holding Problem (GHP)**.
- <u>Conservative vs. Liberal Policies:</u> more ground holding vs. less ground holding (more airborne holding)



Flights scheduled to arrive = 30; Capacity (AAR) = 20

Stochastic Ground Holding Models

- Andreatta, G., and Romanin-Jacur, G. (1987), “Aircraft Flow Management Under Congestion,” *Transportation Science*, **21**, 249-253.
- Richetta, O. and Odoni, A.R. (1993), “Solving Optimally the Static Ground-Holding Policy Problem in Air Traffic Control,” *Transportation Science*, **27**, 228-238.
- Ball, M., Hoffman, R., Odoni, A., and Rifkin, R. (1999), “The Static Stochastic Ground Holding Problem with Aggregate Demands,” Technical Report RR-99-1, NEXTOR, UC Berkeley.

Quote from Ball et al

"Probabilistic information about the uncertain capacity is available in the form of Q scenarios, M_q, for $1 \leq q \leq Q$, where $M_{q,t}$,

$1 \leq t \leq T$, is the arrival capacity of the airport during time t, if scenario q is realized. The probability of the qth scenario occurring is p_q."

Hoffman-Rifkin (H-R) Static Stochastic Ground Holding Model

- Determines number of flights to delay on the ground and number expected to be air delayed per unit time
- Explicitly takes into account the uncertainty of weather
- <u>Formulation:</u>

$$\text{Min} \quad \Sigma_{t=1}\, c_g G_t + \Sigma_{t=1}\, c_a p_q W_{q,t}$$

$$\text{subject to} \quad A_t - G_{t-1} + G_t = D_t \qquad t = 1, \ldots, T+1$$

$$G_0 = G_{T+1} = 0$$

$$-W_{q,t-1} + W_{q,t} - A_t \geq -M_{q,t} \qquad t = 1, \ldots, T+1$$

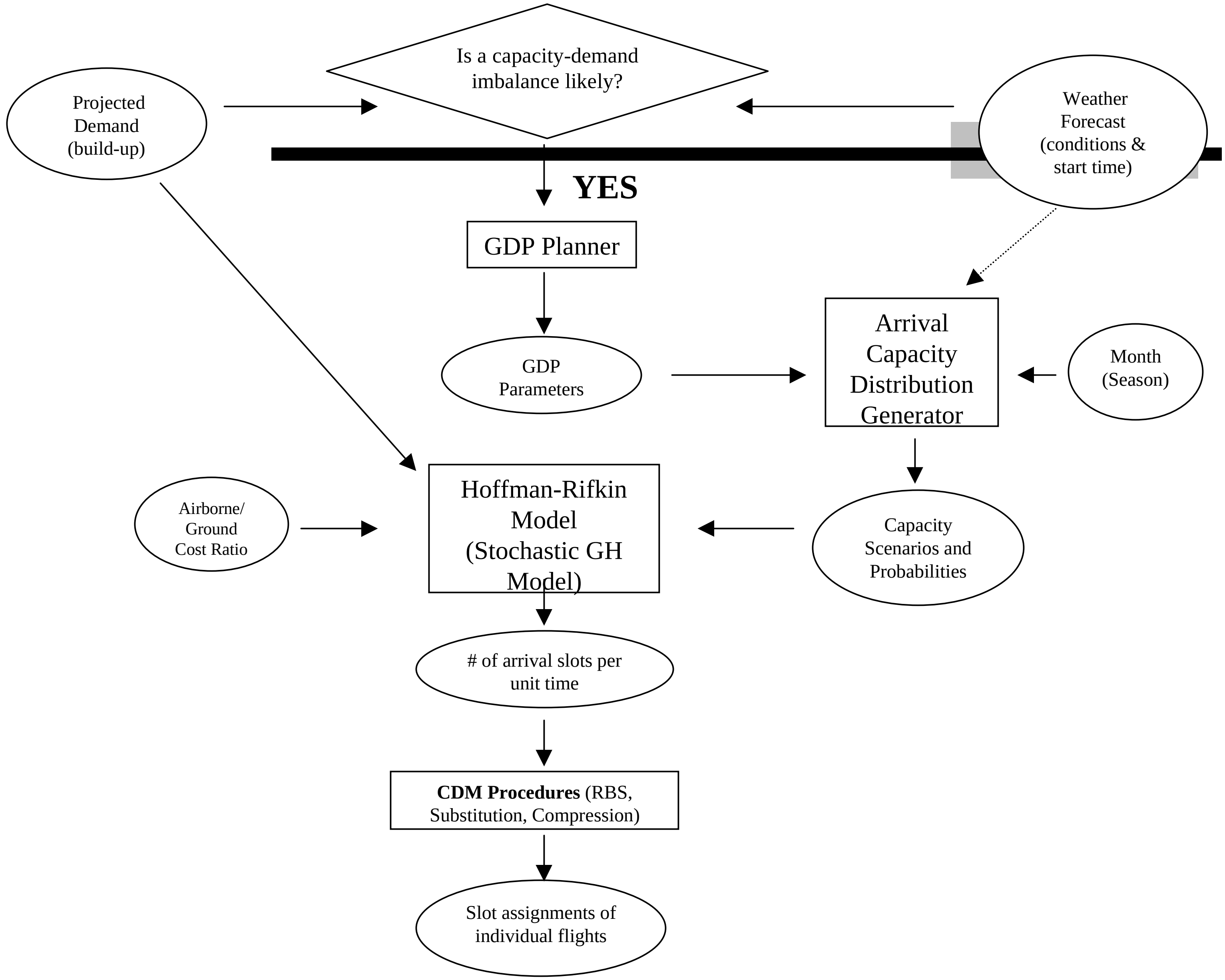
$$q = 1, \ldots, Q$$

$$W_{q,0} = W_{q,T+1} = 0$$

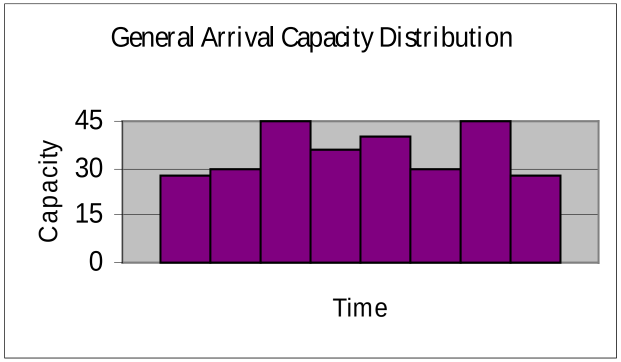
$$A_t \in Z_+,\ W_{q,t} \in Z_+,\ G_t \in Z_+$$

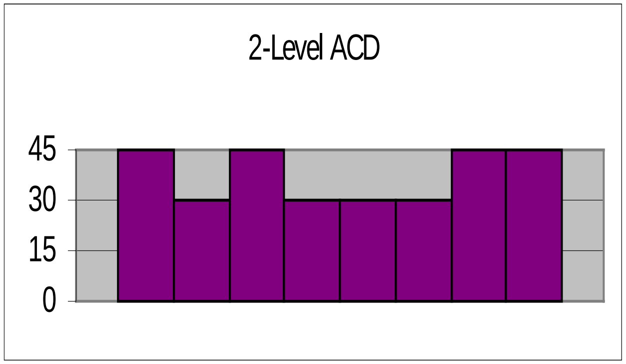
- <u>Inputs:</u> aggregate demand for each time period (D_t), ground delay cost factor (c_g), airborne delay cost factor (c_a), capacity scenarios (Q) and associated probabilities (p_q)

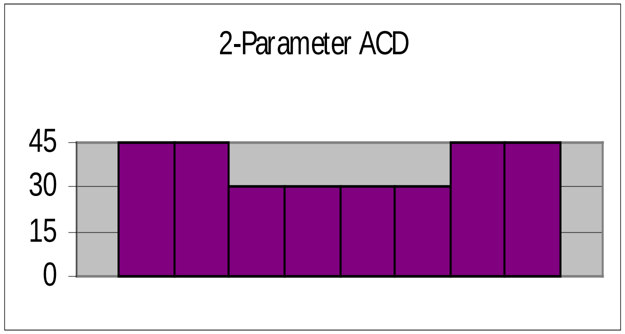
Proposed GDP-E Concept of Operations

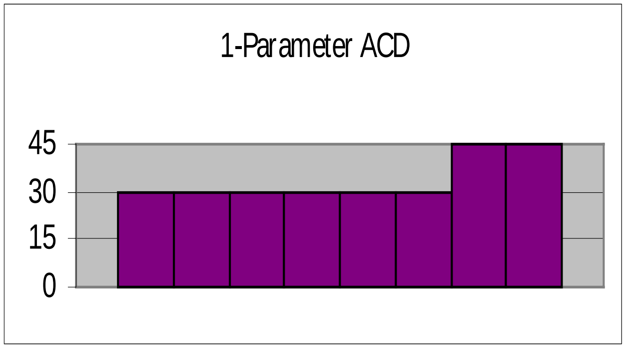


Representative Structures of Capacity Scenarios









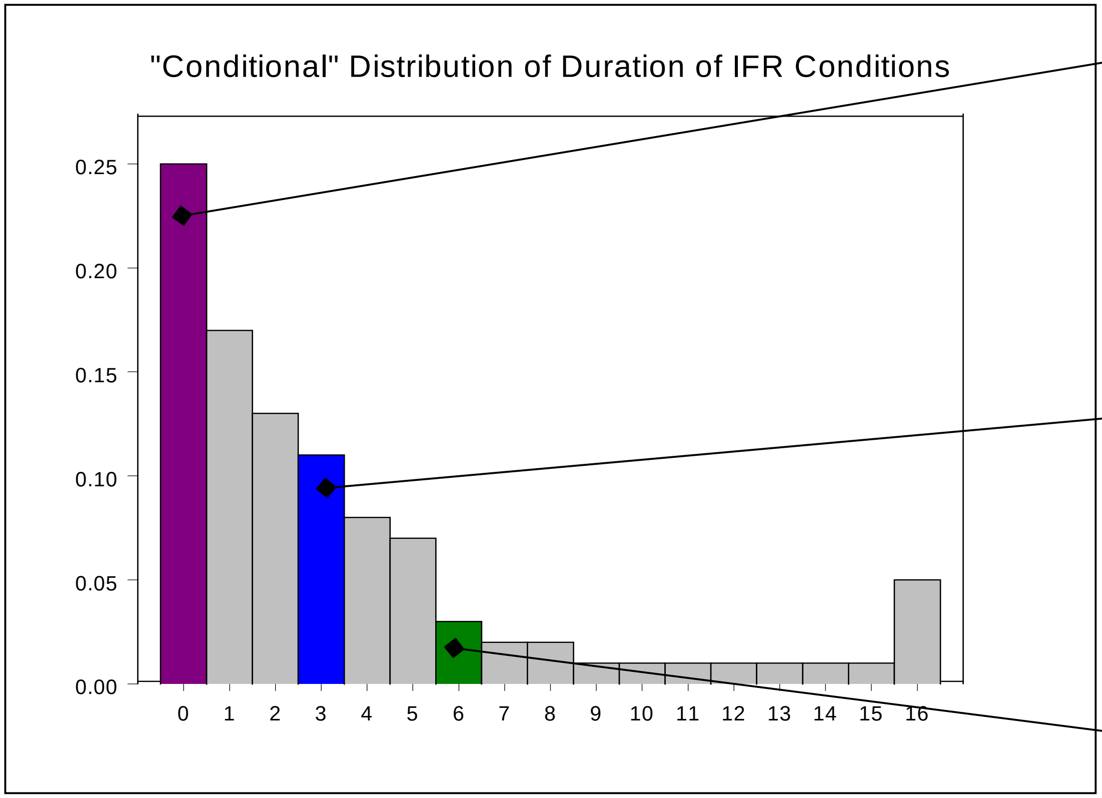
Empirical (Historical) Data Sets

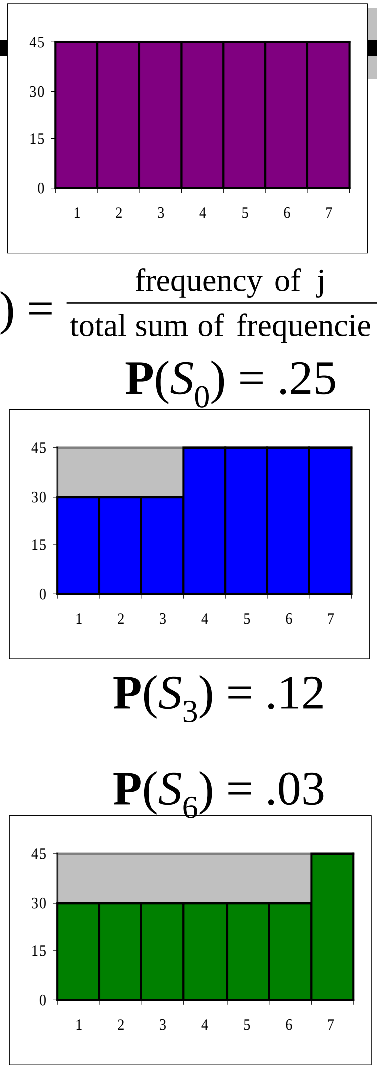
- Ground Delay Programs' Data
  - Logged at ATCSCC and archived by Metron,Inc.
  - Contains GDP parameters such as duration of GDP, scope of GDP and Airport Acceptance Rate (AAR-capacity)
  - Includes 1995, 1996, 1997 GDPs at SFO
  - Can be used for performance analysis
  - Can be used to generate Capacity Probabilistic Distributions Functions (CPDFs) when weather data not available

Data (continued)

- Weather Data
  - Contained in "Surface Airways Hourly" collected by National Climatic Data Center (NCDC)
  - Contains data such as cloud ceiling height, visibility, wind direction and wind speed
  - Can be used to estimate distribution of inclement weather conditions (Instrument Flight Rules-IFR)
- Want combination of GDP data and weather data to get distribution of IFR conditions given a GDP is planned (conditional distribution)

Overall Capacity PDF with 1-Parameter ACDs





$$\mathbf{P}(S_j) = \frac{\text{frequency of } j}{\text{total sum of frequencies}}$$

$\mathbf{P}(S_0) = .25$

$\mathbf{P}(S_3) = .12$

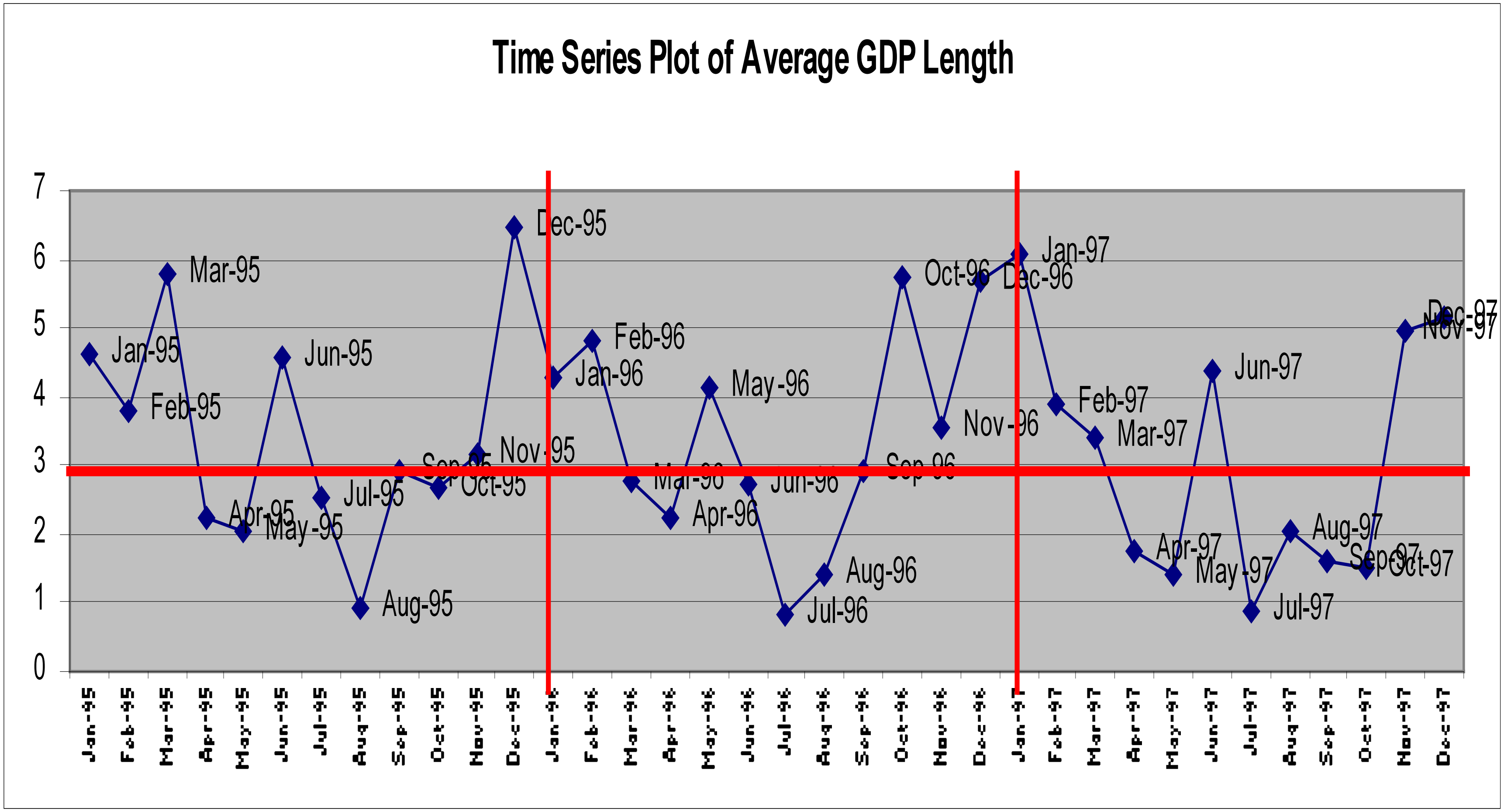
$\mathbf{P}(S_6) = .03$

Relative frequency histogram created by binning historical weather data for San Francisco

Motivation for Deriving Seasonal Distributions

- Weather is a continuous process
- Want to assign (consecutive) months with similar weather to the same season
- Desire to determine a small number of seasons that are operationally efficient

Time Series Plot of Average GDP Length



Varying Distributions by Season

- Determine weather/GDP seasons:
  - Enumerate candidate seasons
  - Season characterized by start and end month (months must be contiguous).
  - Enumerate seasons by different lengths of (contiguous) months.
  - If all possible lengths allowed: 12*11+1 = 133 possible seasons.
  - If length of season restricted to be ≤ 5 months: 12*5 = 60 possible seasons.

Varying Distributions (cont.)

	M_1	M_2	…	M_{12}	M_{13}	M_{14}	…	M_{24}	…	M_{60}
Jan	1	0		0	1	0		1		1
Feb	0	1		0	1	1		0		0
Mar	0	0		0	0	1		0		0
Apr	0	0		0	0	0		0		0
May	0	0		0	0	0		0		0
Jun	0	0		0	0	0		0		0
Jul	0	0		0	0	0		0		0
Aug	0	0		0	0	0		0		0
Sep	0	0		0	0	0		0		1
Oct	0	0		0	0	0		0		1
Nov	0	0		0	0	0		0		1
Dec	0	0		1	0	0		1		1
	x_1	x_2		x_{12}	x_{13}	x_{14}		x_{24}		x_{60}

Choosing Seasons of Least Cost

- Use set partitioning integer program whose formulation is:

  Minimize $\sum_{j=1}^{n} C_j \, x_j$

  subject to $\sum_{j=1}^{n} x_j \leq N$

  $\sum_{j=1}^{n} a_{ij} \, x_j = 1$ for each month i

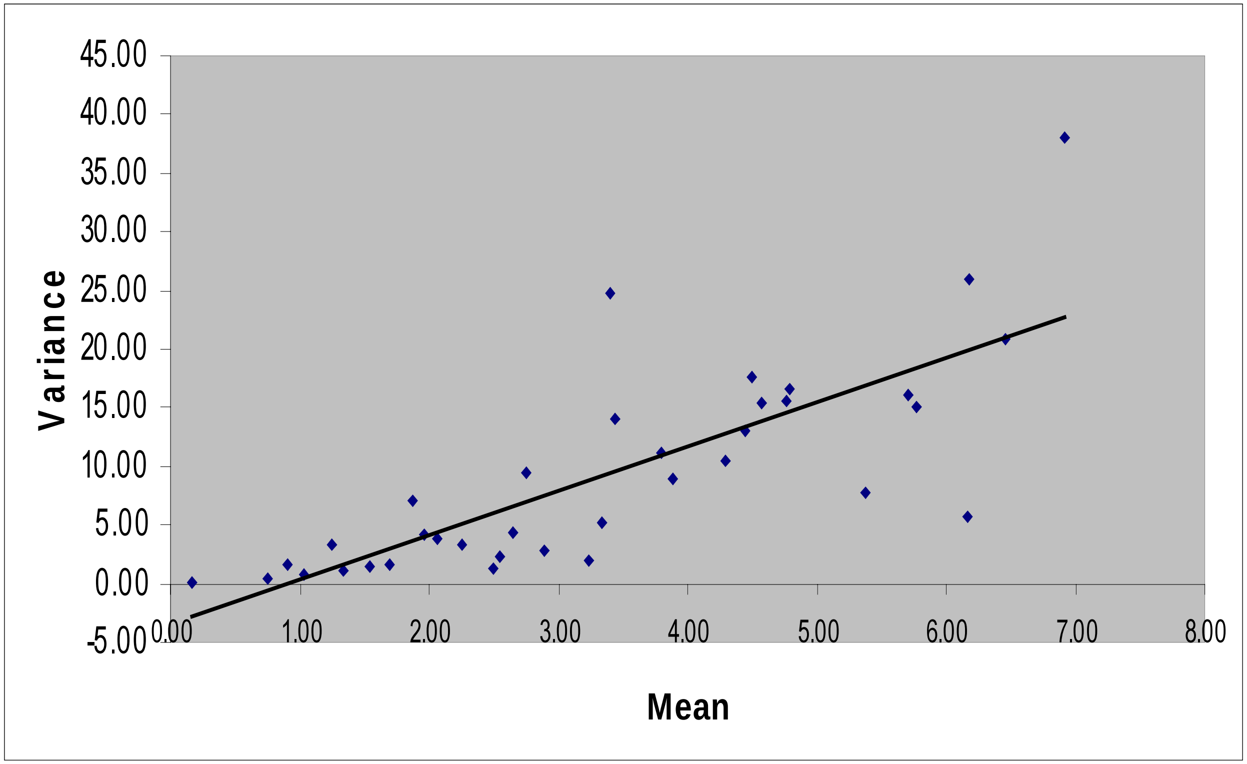
  $x_j \in \{0,1\}$

- C_j is the "cost" of season M_j;
- N is the maximum number of seasons;
- n is the size of the set of candidate seasons.

Costs Based on Differences in Means

Want cost to be based on differences in distributions

(2 parameters, mean and variance)



Cost Functions for Set Partitioning (Differences in Means)

Sum of Squared Deviations (SoSqs)	$\sum_{j=1}^{m}(\overline{X}_{.j} - \overline{X}_{..})^2$
Normalized SoSqs	$\frac{1}{m}\sum_{j=1}^{m}(\overline{X}_{.j} - \overline{X}_{..})^2$
Seasonal Variances	$\frac{1}{m-1}\sum_{j}\sum_{i}(X_{ij} - \overline{X}_{..})^2$

$\overline{X}_{.j}$ is the average over all days *i* in month *j*;

$\overline{X}_{..}$ is the (overall) seasonal average over all days *i* and all months *j*;

X_{ij} is the GDP length on day *i* in month *j*.

Cost Functions Based on Differences in EDFs

- Calculate an EDF for each month *j* (F_j) in a given season.
- Calculate a seasonal EDF (pooled EDF):

$$F = \frac{1}{n}\sum_j (n_j F_j)$$

- Compute the cost of a given season by calculating a Kolmogorov-Smirnov (KS) statistic for the season:

$$KS = \max_x \sqrt{\sum_j \left(\frac{n_j}{n}\right)\left[F_j(x) - F(x)\right]^2}$$

Observations from Computational Experiments

- Different cost functions and max number of seasons yielded different solutions
- Objective functions only include within season interaction and not between season interaction
- Some results may not be operationally feasible (e.g. 3 seasons of length 1, such as results of seasonal variance cost function)

Post Analysis for Evaluating Sets of Seasons

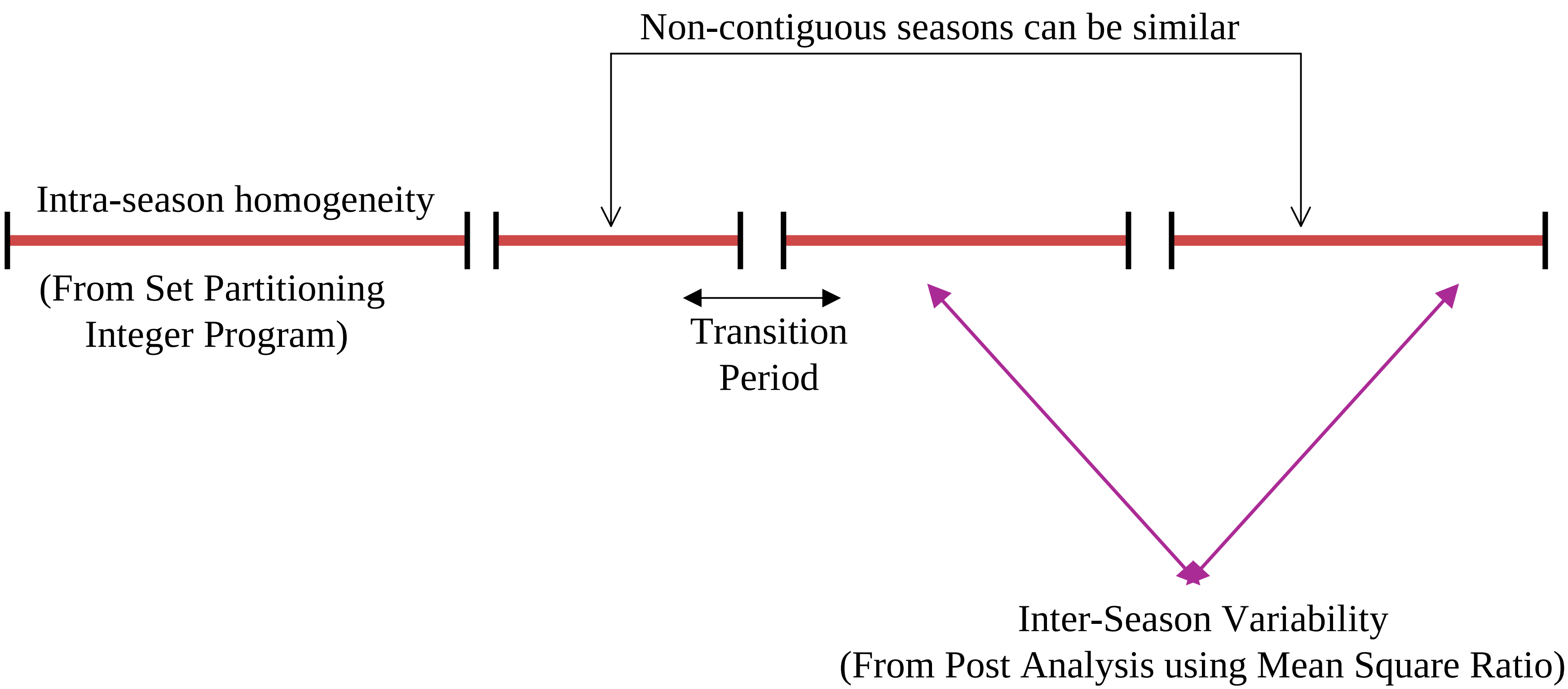
- Single-Factor ANOVA with multiple comparisons

  $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, i=1,…12 and j=1,2,3, $\varepsilon_{ij} \sim N(0, \sigma^2)$

  - Single-factor ANOVA used to test if there exist statistically significant differences in means of seasons.
  - Multiple comparisons used to test for equality between two seasonal means.

- Mean Square Ratio: $$\frac{\left(\dfrac{\sum_s n_s(\overline{Y}_{.s} - \overline{Y}_{..})^2}{k-1}\right)}{\left(\dfrac{\sum_s \sum_j (Y_{js} - \overline{Y}_{.s})^2}{n-k}\right)}$$

Perspectives on Seasonal “Clustering”



Developed in dissertation and used to find seasonal distributions

Results of Post Analysis

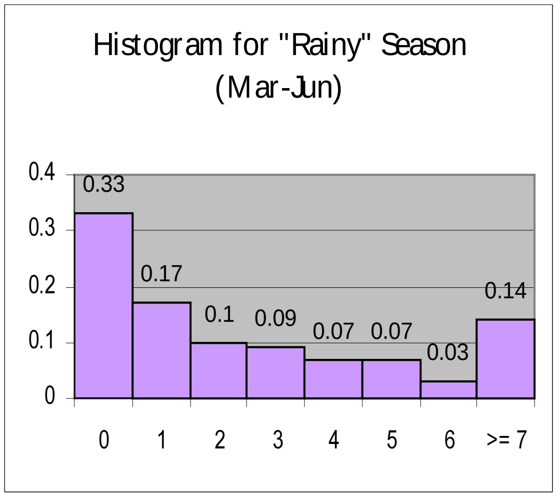
- ANOVA multiple comparisons' results (for seasons resulting from set partitioning of GDP data)

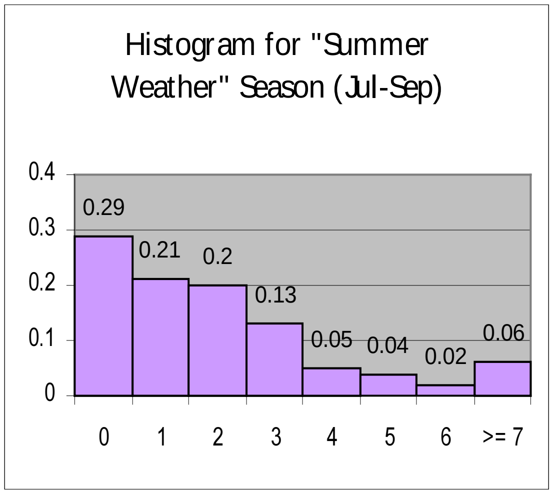
Contiguous Seasons	**P-values**
Apr-Jun vs Jul/Aug	.0288
Jul/Aug vs Sep/Oct	.0388
Sep/Oct vs Nov-Mar	.0053

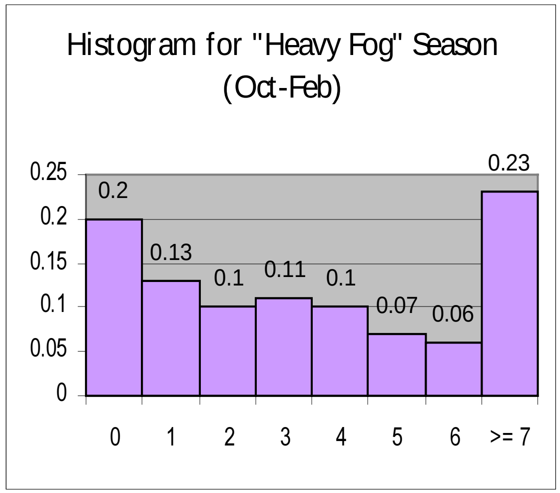
- Mean Square ratios (for seasons resulting from set partitioning of weather data)

Contiguous Seasons	**Mean Square Ratio**
Mar-Jun vs Jul-Sep	14.06
Jul-Sep vs Oct-Feb	24.39

Relative Frequency Histograms for Weather Seasons







Results of H-R Model for Weather Seasons

Weather Season	air delay cost = 1.5	air delay cost = 2.0	air delay cost = 2.5
Mar-Jun (Rainy Season)	2 hours	3 hours	4 hours
Jul-Sep (Summer Weather Season)	2 hours	2 hours	3 hours
Oct-Feb (Heavy Fog Season)	3 hours	4 hours	5 hours

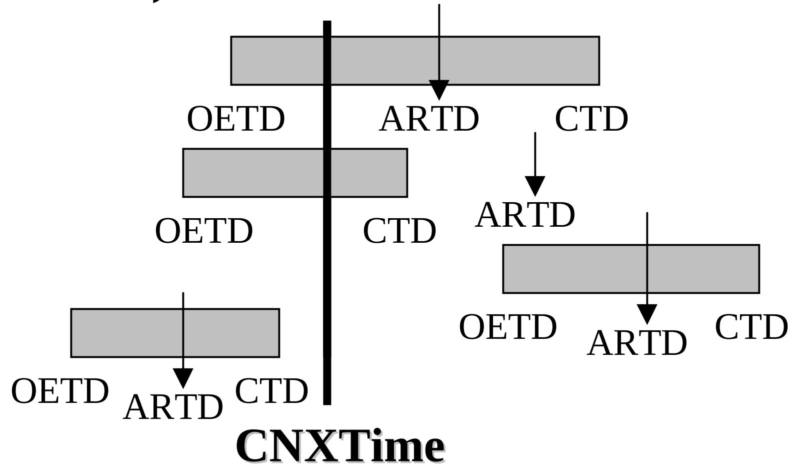
- Air delay costs based on study by Air Transport Association
- Day to day demand in a given season are so similar that results are the same

Observations/Limitations of H-R Model

- Empirically observed that output scenario corresponds to one of input scenarios
- Model does not capture decision-making dynamics
  - Assumes GD is deterministic (independent of capacity scenario realized)
  - Overestimates Airborne Delay (AD)

Adjusting Assigned GD In Canceled GDPs

In a Canceled GDP, some GD is recoverable:



Percentage GD Recovered =

$$1-\left[\frac{ARTD-\max(OETD,CNXTime)}{\min(CTD-OETD,CTD-CNXTime)}\right]$$

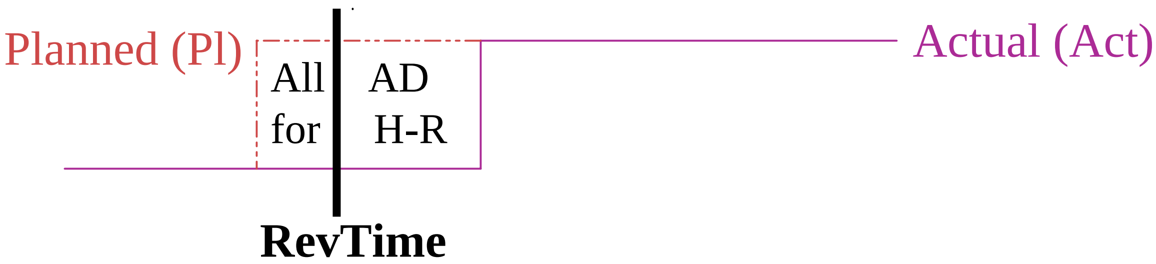
GD Recovered in Canceled GDP

Planned Departure Time – GDP Cancellation Time	% GD Recovered
0-30 minutes	**0 %**
31-60 minutes	40.80 %
61-90 minutes	65.20 %
91-120 minutes	77.15 %
> 120 minutes	**100 %**

Recoverable GD Realized = Assigned Recoverable Delay –
(% GD Recovered)*(Assigned Recoverable Delay)

Adjusting GD in Revised GDPs

In a revised/extended GDP, additional delay incurred (either GD and AD):



If CTD < RevTime, then

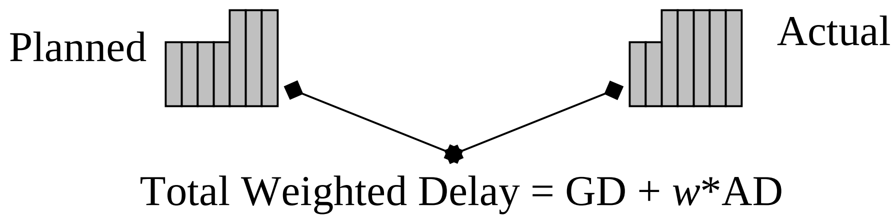
$$CTA_{Act} – CTA_{Pl} = AD.$$

If CTD > RevTime, then additional GD,

$$CTA_{Act} – CTA_{Pl} = GD.$$

Comparison of H-R Results and Command Center Plans

- *Introduction:* Algorithm for comparing planned scenario to actual scenario



  - Used to compare results of model to CC plans
  - Forms basis of new approach (general decision model)
- *Basic Approach:* Order flights sequentially in time, assign to each a new arrival(departure) time and iteratively make decisions
- *Numerical Results*

Numerical Results (M_PAAR vs CC_PAAR)

	Modified H-R	Command Center (ATCSCC)	"Ideal" Plan
Average GD	7284	8914	6875
Average AD	2417	1314	0
Average Weighted Delay	9007	9850	6875

Conclusions/Future Work

- Demonstrated that ACDs used with stochastic models (adjusting delay appropriately) improve the quality of (dynamic) GDPs

Future Work:

- Determine seasonal distributions with arbitrary start and end days
- Formally prove that output capacity scenario corresponds to one of input scenarios
- Model airports using 2-Parameter ACD